ANNUAL REVIEW OF GENETICS

VOLUME 23, 1989

ALLAN CAMPBELL, *Editor*
Stanford University, Stanford

BRUCE S. BAKER, *Associate Editor*
Stanford University, Stanford

IRA HERSKOWITZ, *Associate Editor*
The University of California, San Francisco

ANNUAL REVIEWS INC. 4139 EL CAMINO WAY PO BOX 10139 PALO ALTO, CALIFORNIA 94303-0897


ANNUAL REVIEWS INC.
Palo Alto, California, USA

International Standard Serial Number: 0066-4197
International Standard Book Number: 0-8243-1223-6
Library of Congress Catalog Card Number: 63-8847

♾ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Typesetting by Kachina Typesetting Inc., Tempe, Arizona; John Olson, President
Typesetting Coordinator, Janis Hoffman

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

Annu. Rev. Genet. 1989. 23:289 310

GENETIC ANALYSIS OF PROTEIN STABILITY AND FUNCTION

Andrew A. Pakula

Division of Biology, California Institute of Technology, Pasadena, California 91125

Robert T. Sauer

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

CONTENTS

INTRODUCTION

There is currently a great deal of interest in understanding the amino-acid sequence determinants of protein stability and function. This is important not

0066-4197/89/1215-0289$02.00

only for ongoing studies aimed at dissecting the structure and activities of biologically important proteins, but also for the realization of longer term goals such as the prediction of protein structure from sequence and the design of proteins with novel activities. Detailed genetic and biophysical studies of proteins are beginning to improve our overall understanding of protein structure-function relationships and should allow considerably more progress in the near future.

Genetic studies of protein structure and activity generally center on the properties of proteins altered by deletions or point mutations. Two basic strategies are commonly used. In the first, one creates a specific alteration in the coding sequence and asks, "What is the effect of this alteration?" In the second, one creates pools of randomly altered genes, applies a screen or selection to identify those encoding proteins with a specific phenotype, and then asks, "What kinds of sequence alterations can cause this effect?" The directed approach is most useful when there is already enough information about the structure or activity of the protein to formulate specific questions about the roles of particular residues. The random approach is particularly useful for identifying important residues in an unbiased way in the absence of detailed information from other sources or studies.

MAKING AND MAPPING MUTATIONS

Traditionally, mutations have been generated by treating cells with agents such as nitrosoguanidine, EMS, and UV light. These mutations are then located by genetic mapping. For the study of protein function and stability, this approach is rapidly being replaced by methods involving manipulations of cloned genes. Mutations may be generated by directed mutagenesis, rapidly localized to specific restriction fragments using recombination in vitro, and then analyzed by DNA sequencing.

Numerous methods are available for the random mutagenesis of cloned genes. In general, these permit a broader, less biased, mutagenic specificity than has been possible with traditional techniques. Furthermore, several strategies are available for limiting random mutagenesis to portions of a DNA molecule. Thus, a specific gene, or only selected regions of a gene, can be mutagenized without creating changes in the rest of the cloning vector. Specific mutations can be constructed in cloned genes using oligonucleotide-directed mutagenesis. This technique permits the creation of proteins with one or more defined amino acid change(s). Such changes can also be created by synthesis of double-stranded DNA cassettes that are then returned to the gene in vitro. Cassette mutagenesis can also be used as a powerful technique for localized random mutagenesis when some or all of the base positions in the cassette are synthesized with a mixture of wild-type and mutant nucleotides.

Thus, single nucleotide pairs, single codons, or blocks of codons may be randomly mutagenized with extremely high efficiency. Many such methods are described in several reviews (7, 66, 79) and are not discussed here. Instead, we concentrate on general mutagenic strategies and discuss what these methods have taught us regarding the sequence determinants of protein structure, stability, and activity.

The domain is the basic unit of protein structure and function. For proteins with multiple domains, deletion analysis can often rapidly identify large portions of a protein sequence that are not required for a particular activity. For example, in *Escherichia coli* alanine-tRNA synthetase (875 residues), the COOH-terminal 415 residues of the protein can be deleted and the truncated protein still retains amino acylation activity (27). In like fashion, the site-specific DNA-binding activities of the GAL4 (881 residues) and GCN4 (281 residues) transcriptional regulatory proteins of *Saccharomyces cerevisiae* reside in independent domains of 60–100 residues (26, 31). In such cases, it clearly makes sense to use deletions to identify structural domains and thus restrict the problem being investigated to the greatest possible extent. However, the effects of deletions (or insertions) within structural domains are generally too drastic to provide very much useful information for structure-function studies. At this more detailed level, missense mutations provide the major tool for further dissection of structure and activity.

Studies of mutant proteins can be roughly divided into two classes: Some focus on the identification of residues that are directly involved in binding or enzymatic activities. Others concentrate on the importance of specific residues and interactions in the folding of proteins, and in the stability of protein structures. Although these two types of studies have clearly different goals, they are intimately related in the sense that protein folding and the maintenance of a stably folded structure are almost always prerequisites for activity. Thus, putative active-site mutations must be shown to be free of severe effects on structure and stability, and putative stability mutations must be distinguished from those that disrupt function but not structure.

FOLDED AND UNFOLDED PROTEIN STRUCTURES

As a rule, proteins fold and unfold spontaneously in a reaction that can be described in terms of a simple, two-state equilibrium. The unfolding of a monomeric protein can be modeled as

$$\text{folded} \rightleftharpoons \text{unfolded} \qquad K_u = \frac{[\text{unfolded}]}{[\text{folded}]}$$

where the equilibrium constant, K_u, is a measure of the ratio of unfolded to

folded protein molecules. The free energy change upon unfolding (ΔG_u) can be calculated from K_u by

$$\Delta G_u = -RT \ln (K_u)$$

where ΔG_u represents the difference between the free energies of the folded and the unfolded states, R is a constant (1.98 cal/mol-°K), and T is the temperature in °K. The conversion from terms of the equilibrium constant to terms of free energy is useful, because this permits the net stability of the folded protein to be directly compared to the energetic contributions of specific interactions. At 37°C, a 1 kcal/mol decrease in ΔG_u corresponds to a fivefold increase in K_u. Values of ΔG_u for protein unfolding range from about 3–15 kcal/mol under physiological conditions of temperature and pH (49). Single destabilizing mutations can decrease the stability of some proteins to the point where most molecules are unfolded (see discussion on Destabilizing Mutations). If ΔG_u for a protein is 3 kcal/mol at 37°C, for example, then the fraction of unfolded protein would be 0.7%. A mutation that decreased the stability of this protein by 4 kcal/mol would increase the fraction of unfolded protein to 80%. Hence, a fivefold loss in activity would be expected simply on the basis of the decreased concentration of folded, active molecules. In reality, the activity loss could be considerably greater if, for example, the destabilizing mutation also affected the specific activity of the folded protein. Moreover, in the cell, processes such as aggregation or proteolysis that rapidly and irreversibly remove unfolded protein, may magnify the phenotypic effects of destabilization.

Protein structures contain an impressive array of stabilizing interactions; these include hydrophobic and packing interactions, hydrogen bonds, and salt bridges. As a result, it is often difficult to imagine that changing a single side chain could result in a serious perturbation of structure or stability. However, although the forces favoring protein folding contribute a large amount of energy and involve a large number and variety of interactions, they are nearly offset by the entropic cost of folding. This entropic penalty is due to the enormous loss of conformational freedom that occurs as the protein goes from a denatured state with many possible conformations to a native state with only one or a few conformations. Thus, a net stability of 5 kcal/mol may arise as the difference between a favorable energy of 300 kcal/mol and an unfavorable energy of 295 kcal/mol. Clearly, in such a case, small fractional changes in the energies favoring and opposing folding can shift the balance and lead to unfolding. Such changes can occur as a consequence of alterations in temperature, pH, and the concentration of denaturants, as well as by the introduction of mutations.

Measuring Protein Stability

The folded and unfolded forms of a protein almost always have different spectral or hydrodynamic properties. As a result, the fraction of unfolded protein molecules can generally be determined by monitoring an appropriate physical property as a function of changes in temperature, urea concentration, or guanidinium-HCl concentration (49). Susceptibility to proteolysis provides another means of determining protein stability, because most native proteins are relatively resistant to cleavage, whereas denatured proteins are exquisitely sensitive (50). Thus, the rate at which a purified protein is degraded by a protease will depend on the fraction of molecules that are unfolded, and proteolysis in vitro can be used to compare the stabilities of a mutant protein and its wild-type counterpart (22).

Most single-domain proteins unfold in a cooperative fashion, i.e. a given molecule is either folded or unfolded. Figure 1 shows thermal denaturation experiments for a wild-type protein and a mutant that displays reduced stability. Although the fraction of molecules that are unfolded varies as a function of temperature for both proteins, the unfolding transition for the mutant occurs over a lower temperature range than that for the wild-type protein. At any given temperature in the transition zone, K_u and ΔG_u values can be calculated for both molecules. The difference in stability can then be expressed as $\Delta\Delta G_u$, which is defined as ΔG_u(wild-type) − ΔG_u(mutant). We refer to $\Delta\Delta G_u$ values when we say, for example, that an Ile→Val substitution destabilizes a protein by 1 kcal/mol. In many cases, it is also convenient to refer to the temperature at which half of the protein molecules are unfolded, T_m, as a rough measure of the stability of a protein.

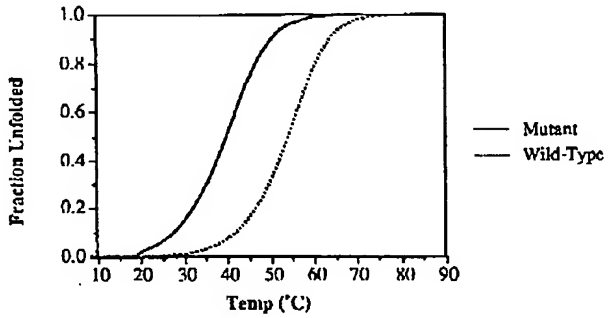

Figure 1 Thermal denaturation of a hypothetical protein and a mutant derivative with reduced stability. The stabilities of the two proteins can be directly compared at temperatures where their transition zones overlap (ca. 35–50°C in this case). Stabilities at temperatures outside of the transition zone can be calculated if ΔH and ΔC_p for unfolding are known (for discussion, see references 6, 49).

TOLERANCE OF RESIDUE POSITIONS TO SUBSTITUTIONS

An initial question concerning any protein is how many of its amino acids are really critical for structure or function? Can the protein be inactivated by substitutions anywhere in the sequence or are only a few key residues really important? In the sections that follow, we first discuss studies of defective mutations for several proteins with known three-dimensional structures and for which deleterious mutations have been isolated and identified by random mutagenesis. Subsequently, we discuss the phenotypically neutral mutations that have been studied in several proteins.

Mutations Causing Reduction or Loss of Activity

Many different missense mutations, each causing a defective phenotype, have been isolated in the genes encoding staphylococcal nuclease (67), phage T4 lysozyme (2), the N-terminal domain of λ repressor (20), λ Cro (52), and yeast iso-1-cytochrome *c* (18, 19). The severity of the mutant phenotypes varies for the different proteins, and often varies among the collection of missense alleles for a given protein. For example, the T4 lysozyme mutants were each isolated on the basis of a temperature-sensitive phenotype, and thus known to be able to fold and function at the permissive temperature. By contrast, many of the mutant forms of the other proteins showed no activity at any temperature. Nevertheless, in each case, defective mutations can clearly occur at many positions. For example, residue substitutions at 32 of the 66 positions in λ Cro, and 55 of the 149 positions in staphylococcal nuclease are known to result in diminished activity or loss of activity. Moreover, the sites of these mutations are not obviously clustered in the protein sequences or within the crystal structures of any of the five proteins.

What kinds of mutations result in a defective phenotype? The striking observation is that most mutant substitutions appear to affect activity indirectly via effects on protein structure or stability. This conclusion is supported by several findings. First, these mutations occur at positions for which no evidence exists for a direct functional role; that is, they are found at sites distant from the active site/binding regions of the proteins. Second, several mutant proteins of this class have been purified and shown to be less stable than wild-type for each of the five proteins (6, 22, 51, 55, 62, 68). Finally, most of these mutations affect side chains that would be expected to play important structural roles. These include side chains that are buried in the protein structure, side chains involved in hydrogen bonds or electrostatic interactions, and side chains with special properties, such as glycine and proline. We return later to a discussion of each of these types of mutations.

The degree to which a side chain is buried in the native protein is usually

defined by computer calculation of its fractional accessibility to water (57); low solvent accessibilities indicate that residues are buried, whereas high accessibilities indicate that residues are exposed on the protein surface. In Figure 2 the likelihood of isolating a destabilizing substitution is plotted as a function of the fractional accessibility of the wild-type side chain for staphylococcal nuclease, T4 lysozyme, λ repressor, λ Cro, and yeast iso-1-cytochrome *c*. Buried or core residues are obviously the most common sites of destabilizing mutations for each of the five proteins, suggesting that these residues are extremely important for the maintenance of protein structure and stability. However, certain exposed or partially exposed side chains must also be structurally important, as some destabilizing mutations also occur at these positions.

As might be expected, at least some of the mutations that disrupt protein activity do so in a direct fashion. In staphylococcal nuclease, λ repressor and λ Cro, approximately one quarter to one third of the defective mutations alter residues that are directly involved in function. In staphylococcal nuclease, these include substitutions at twelve positions within the active site or polynucleotide binding region. In λ repressor and λ Cro, mutations occur at about ten positions that form a significant portion of the DNA-binding surfaces of each protein. Active site mutations are not represented among the T4-lysozyme mutants, but are not expected as these temperature-sensitive mutants have wild-type or near wild-type activities at low temperatures. None

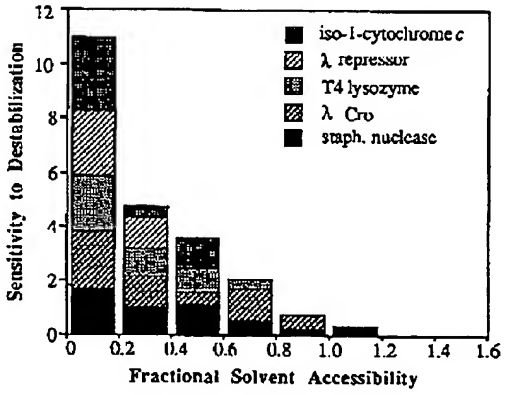

Figure 2 Normalized probability of isolating destabilizing mutations as a function of fractional residue solvent accessibility. Data are compiled for the N-terminal domain of λ repressor (20), phage T4 lysozyme (2), λ Cro (52), staphylococcal nuclease (67), and yeast iso-1-cytochrome *c* (19). For each accessibility class, the number of positions at which destabilizing mutations occur was divided by the total number of positions in that class. This value was then normalized by dividing by the frequency of such mutations for the entire protein. Thus, a value of 1 indicates residues with an average susceptibility to destabilizing mutations.

of the mutations in iso-1-cytochrome *c* abolishes electron transfer function while retaining normal stability. The absence of this class of mutations may simply reflect the sampling of potential mutants or could indicate that functional residues in this protein play a dual role and are involved both in function and in structural stability.

The generalization that most missense mutations act by affecting protein stability does not hold in all cases. For example, missense mutations causing a null phenotype in EcoR1 primarily affect residues located at the protein-DNA interface or at the protein-protein dimer interface (77). Only a few mutations apparently alter the stability of the monomers in this case. It is possible that the EcoR1 monomer is more stable than the proteins discussed above. In this case, mutations that are sufficiently destabilizing to cause a complete loss of activity might be rare. Alternatively, the clustering of mutations in the EcoR1 case may reflect the stringency of the mutant selection, which demands elimination of even trace amounts of enzyme activity.

Neutral Mutations

The studies discussed above show that defective mutations can generally occur at many positions throughout a protein sequence. What, however, can be concluded about the sites where mutations were not isolated? Are amino acid changes at these positions silent or is the catalog of defective mutations simply incomplete? A general way of addressing this question is to ask whether amino acid substitutions can be functionally tolerated at any given residue position. Two strategies have been used for the efficient generation of mutations that can then be scored for a neutral phenotype. The first is tRNA-mediated suppression of *amber* (UAG) mutations. Strains have now been isolated or constructed that allow the efficient insertion of Ala, Cys, Gln, Gly, His, Leu, Lys, Phe, Pro, Ser, and Tyr at *amber* codons (for review, see ref. 47). The second involves codon randomization via cassette mutagenesis (56). Here a double-stranded DNA cassette is chemically synthesized with one or more codons randomized by the inclusion of all four bases during synthesis. The cassette is then recloned into the gene and introduced into cells by transformation. Genes encoding active proteins can then be identified by a selection or screen, and sequenced.

Miller and colleagues have provided the most extensive view of neutral mutations in their work on suppressed nonsense mutations at 142 of the 360 codon positions in the *lac* repressor gene (35, 44). In these studies, the phenotypes of some 1500 single residue changes were scored and approximately half of these changes were found to be phenotypically silent. At 28 residue positions, all substitutions tested were tolerated. At an additional 54 sites, at least half of all substitutions were tolerated and, in many of these cases, only proline was not tolerated. These results indicate that the identity of

the side chain is not a critical determinant of either structure or activity for a significant number of positions in a protein.

Figure 3 shows the nearly exhaustive set of neutral mutations that have been isolated in part of the N-terminal domain of λ repressor following cassette mutagenesis (56; J. Reidhaar-Olson, unpublished data). All the residues in the mutagenized region are distant from the DNA in the crystal structure of the protein-DNA complex (28), and thus any effects on activity caused by substitutions in this region must be mediated indirectly via protein structure or stability. As shown in the figure, the neutral mutations isolated at six positions included only the wild-type residue or a single conservative substitute. The method of mutagenesis used in these experiments ensures that all residue substitutions are represented in the population prior to selection. Hence, recovery of only a few neutral substitutions indicates that most other substitutions are not functional and have been selected against. Five of these positions are buried in the active dimeric form of N-terminal domain. At most surface positions, however, a large number of chemically different side chains are allowed, including those that are charged, uncharged, large, small, hydrophilic, and hydrophobic. Clearly, residue positions that display this degree of tolerance do not play essential roles in protein structure or stability. By contrast, the finding that allowed substitutions are highly restricted for buried residues suggests that these side chains carry fundamentally important information for protein folding and stability.

DESTABILIZING MUTATIONS

Substitutions Affecting the Hydrophobic Core

It should be evident from the preceding sections that residues buried within the core appear to be extremely important determinants of protein structure and stability. In proteins of known structure, the cores are composed chiefly of hydrophobic residues and, more rarely, of polar residues that can satisfy

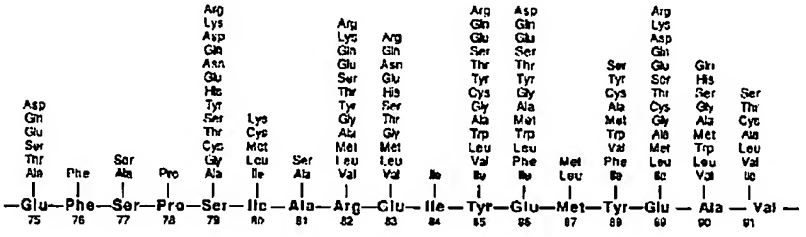

Figure 3 Neutral residue substitutions in the N-terminal domain of λ repressor. [Data for residues 84–91 are from ref. (56). Data for residues 75–83 are from the unpublished work of John Reidhaar-Olson.]

their hydrogen-bonding potential by forming hydrogen bonds with the protein backbone or other side chains (13). These core residues pack together efficiently to fill the protein interior (57). The characteristics of close packing and hydrophobicity are presumably important for two reasons. First, these internal packing interactions must, in some sense, determine the overall shape of the protein. Second, because of the hydrophobic effect, the shielding of nonpolar-core side chains from water contributes to the stabilization of the native protein structure (30).

In an unfolded protein, water is thought to be organized in cage-like structures around the hydrophobic side chains. When these side chains are transferred to the nonpolar environment of the core, the structured water is released. This increases the entropy of the solvent and thereby helps to stabilize the native protein. The magnitude of the hydrophobic effect may be estimated from free energies calculated for the transfer of amino acid side chains from water to nonpolar solvents such as ethanol or octanol (14, 48). These calculations suggest that the hydrophobic effect provides the largest free energy contribution to the stability of folded proteins (13, 30).

The hydrophobic contribution of individual buried side chains to stabilization has been examined in a number of different proteins (33, 40, 41, 53, 78). For each, the effects on protein stability of several different substitutions at a single site have been determined. In T4 lysozyme, for example, the effects of 13 different substitutions for Ile3, a residue that is about 80% buried in the structure, have been determined (40). The most deleterious mutations involve replacing Ile3 with larger side chains such as Trp and Tyr, or polar or charged residues such as Ser, Thr, and Asp. Each of these substitutions decreases stability by 1.7 to 3.2 kcal/mol. By contrast, substitution of Ile3 with the smaller but nonpolar Ala was found to decrease stability only by 0.7 kcal/mol.

It is difficult, however, to draw any general conclusions regarding the expected magnitudes of stability changes resulting from changes at buried or partially buried positions. For example, in the small ribonuclease barnase, substitution of the buried Ile196 side chain by Ala results in a destabilization of 4.0 kcal/mol (33). Although this Ile→Ala change is chemically identical to the Ile→Ala change in lysozyme, the observed destabilization is 5–6 times larger for barnase than for lysozyme. A similarly large destabilization has been observed for the Leu57→Ala mutation in λ repressor (53). This mutation reduces stability by 4–5 kcal/mol and reduces the T_m of the protein from 54°C to 20°C.

Why are the destabilizing effects of Ile→Ala or Leu→Ala mutations so different in the different cases? One possible factor is the degree to which the side chain being studied is truly buried within the protein. For the T4-lysozyme studies, Ile3 is only partially buried and is quite near the protein surface. Hence amino acid substitutions can probably be accommodated by

local adjustments in packing interactions, and polar side-chain atoms can probably satisfy their hydrogen-bonding needs by extending out into solvent. This in fact happens for the Ile3→Tyr change (40). By contrast, the Ile and Leu residues replaced in barnase and λ repressor are completely buried in the hydrophobic core. Replacing these residues with a smaller side chain like Ala may require leaving an energetically unfavorable hole in the hydrophobic core. Here, the reduced stability will result both from the loss of hydrophobic interactions and from the cost of having a cavity in the protein interior (33). Cavities are presumed to be energetically expensive because some van der Waals interactions between the protein and water in the unfolded state will not be replaced by energetically comparable interactions within the folded protein. As a result, there will be a net decrease in stability due to these lost packing interactions.

The reader should not be left with the impression that the most deleterious mutations in the hydrophobic core will only decrease protein stability by 4–5 kcal/mol. In fact, for completely buried positions such as those discussed above, replacing hydrophobic residues such as Leu and Ile with extremely polar or charged residues seems likely to destabilize the protein to a considerably greater extent. Such large changes in ΔG_u are less likely to be measured than more moderate ones owing to the technical difficulties involved in purifying and studying extremely unstable proteins.

Glycine and Proline Substitutions

Glycine lacks a β-carbon and can therefore assume many backbone dihedral angles that are energetically unfavorable for other amino acids (13). This property is extremely important because it allows glycine to be used in certain types of reverse turns where positive dihedral angles are required (59). Replacing glycines in such turns with any other residue would be expected to be destabilizing unless the protein could form an alternative type of turn. In λ Cro, the destabilizing Gly48→Ala and Gly15→Glu substitutions affect glycines with positive dihedral angles in turns, and thus presumably act by this mechanism (52).

The pyrrolidine ring of proline constrains its Φ dihedral angle to values near −60°. Thus, proline should be destabilizing at positions where significantly different backbone torsional angles are required. An example occurs in staphylococcal nuclease, where replacing His121 ($\Phi = -170°$) with Pro results in complete loss of activity (67). In addition, proline is not found in the middle or at the C-terminal ends of most α-helices (12, 58). The exclusion of proline from helices is thought to be a consequence of steric clashes between the pyrrolidine side chain and the β-carbon of the previous residue (64) and/or because one of the α-helical hydrogen bonds is lost as a result of proline not having a peptide-NH group. Destabilizing mutations in which α-helical

residues are replaced by prolines are reasonably common. For example, the Leu12→Pro and Ser35→Pro defective mutations in λ repressor (20) both affect surface residues in α-helices. Clearly, the effects of these mutations are caused by insertion of the proline and not by the loss of the wild-type side chain, because other substitutions such as Leu12→-Gln or Ser35→Leu are fully functional at both positions (21; J. Reidhaar-Olson, unpublished data).

Substitutions of the type Xaa → Gly or Pro → Xaa (where Xaa represents any other amino acid) may cause destabilization by increasing the entropy of unfolding (43). These entropy increases would occur because the backbone of glycine has more accessible conformations in the unfolded state than other residues, whereas the backbone of proline has fewer accessible conformations. The Pro35→Leu and Pro76→Leu mutations in yeast iso-1-cytochrome *c* may destabilize the protein, in part, by this mechanism (19, 55). However, both of these substitutions affect residues that are inaccessible to solvent and thus also alter packing and hydrophobic interactions in the core. Surface mutations of the reverse type, Gly → Xaa and Xaa → Pro result in stabilization of T4 lysozyme (43) and λ repressor (23) but only by about 0.4–0.8 kcal/mol. These free energy changes may represent the degree of destabilization that results solely from the conformational entropy changes that occur upon replacing proline or introducing glycine.

Substitutions Affecting Hydrogen Bonds and Electrostatic Interactions

Several uncertainties make it is difficult a priori to assess the importance of hydrogen bonds or salt bridges in protein structures. First, any hydrogen bond or electrostatic interaction that is made in the folded protein is formed at the cost of breaking similar bonds with solvent in the unfolded form. Second, the strength of electrostatic interactions depends on the extent to which they are shielded by solvent, and it can be difficult to assess these shielding terms for interactions at protein surfaces. Nevertheless, in model systems involving enzyme-substrate binding, the energetic contributions of hydrogen bonds that do not involve charged residues range from 0.5 to 1.5 kcal/mol, whereas hydrogen bonds involving charged residues may contribute as much as 4 kcal/mol (5, 15, 71). The experimental studies described below suggest that hydrogen bonding and electrostatic interactions can contribute modestly to protein stabilization.

A significant number of the destabilizing mutations in staphylococcal nuclease, λ Cro, λ repressor, and T4 lysozyme affect residues whose side chains participate in hydrogen bonds. For example, the Thr157→Ile mutation in T4 lysozyme disrupts a network of hydrogen bonds mediated via the threonine hydroxyl group (17). Studies of the stabilities and structures of a set

of mutants at position 157 suggest that loss of this side chain hydrogen bond decreases the overall stability of T4 lysozyme by about 1.2 kcal/mol (3). A similar degree of destabilization has been measured for the Thr113→Val mutation in dihydrofolate reductase, which disrupts several hydrogen bonds mediated by the wild-type side chain (54).

In proteins, salt bridges can occur between positively and negatively charged side chains. Surface salt bridges in bovine pancreatic trypsin inhibitor (10) and dihydrofolate reductase (54) seem to contribute about 1 kcal/mol to overall stability, although hydrogen bonding may also play a role in the latter case. Electrostatic interactions between charged side chains and the ends of α-helices are also possible; because of the alignment of the peptide dipoles, α-helices bear a partial positive charge at their N-terminal ends and a partial negative charge at their C-terminal ends (25). Stabilizing interactions of this type in T4 lysozyme (46) and barnase (60) appear to contribute from 0.8 to 2 kcal/mol to protein stability.

Substitutions Affecting the Denatured State

As we have seen, it is often possible to rationalize the effects of destabilizing mutations in terms of the folded structure of a protein. Matthews (42) has argued that this suggests that most destabilizing substitutions exert their effects primarily on the folded state of a protein. However, Shortle and his colleagues have found that several mutations in staphylococcal nuclease alter the physical properties of the unfolded state (68–70). Because the overall stability of a protein depends on the free energies of both the folded and unfolded states, it is not unreasonable that a mutation could exert its effect primarily via the unfolded state. However, at present, it is not clear how to partition the effects of the staphylococcal nuclease mutations between perturbation of the energies of the unfolded and folded states.

We have already mentioned substitutions involving proline and glycine that may affect protein stability by altering the conformational entropy of the unfolded state. Disulfide linkages between cysteine residues are also thought to stabilize folded proteins by reducing the number of conformations accessible to the unfolded protein and thus reducing the entropy of unfolding (30). Although disulfide bonds are extremely rare in intracellular proteins, they are common in secreted proteins and provide potential targets for destabilizing mutations. The introduction of new disulfide bonds has been a common strategy for attempting to increase protein stability through rational design. However, the stabilization afforded by such covalent cross-links is highly dependent upon structural context and position. Some new disulfides do stabilize the protein, while others have no effect, or may actually destabilize the protein (73, 74).

CONFORMATIONAL CHANGES IN MUTANT PROTEINS

Until now, we have been discussing amino-acid sequence changes in terms of their effects on the equilibrium between the folded and unfolded conformations of a protein. It is also worth asking if single point mutations can cause significant changes in activity by altering the conformation of the folded protein. There are a few cases where mutations have been shown to cause propagated conformational changes. For example, replacing Pro86 on the surface of T4 lysozyme causes a conformational change by allowing extension of an α-helix (1). Although the observed structural change is modest (residues 81–83 shift positions by no more than 1.4 Å), some of these changes occur 11 Å from the site of the mutant substitution. Another case in which a single substitution causes nonlocal conformational changes occurs in staphylococcal nuclease. Here, a Glu43→Asp substitution at a partially buried position in the enzyme active site results in detectable changes at residues as far as 30 Å away (75). In both the lysozyme and nuclease cases, however, the observed changes are small in terms of the overall structure. Moreover, because proteins are somewhat flexible, it is not obvious a priori that small conformational changes would cause large reductions in activity. The cases discussed do not resolve this issue. In the T4-lysozyme case, the observed changes are distant from the active site, and the mutant enzyme has normal stability and activity. In the staphylococcal nuclease case, the mutation alters an active site residue, and thus it is difficult to determine the extent to which the loss of activity is caused by the conformational change.

Overall, misfolding appears to be rare. Most mutant proteins that have been studied thus far have conformations that are extremely similar to wild type. For example, Matthews and his colleagues have solved the crystal structures of more than 50 mutant forms of T4 lysozyme and found that in almost all cases the mutant and wild-type structures are extremely similar, with structural differences occurring only at or near the site of the mutant substitution (3, 17, 40, 42, 43, 46). This is true even for mutant proteins that are significantly less stable than wild-type.

IDENTIFYING RESIDUES IMPORTANT FOR FUNCTION

Many genetic analyses of proteins are directed towards answering functional questions rather than those concerning protein structure or stability per se. Which are the active site residues? Which residues mediate binding and specificity. These questions have been approached both by studies of defective mutants and by studies of neutral mutations.

As we have seen, mutations affecting active site residues are usually present in collections of defective mutations, but so are mutations that affect

structure and stability. Hence to conclude that a defective mutation affects a functionally important residue, it must first be shown that it does not affect structure or stability. This has been established in some cases by purifying the mutant proteins and determining their stabilities. For example, in studies of defective λ repressor mutants, it was shown that a subset of the mutant proteins had thermal stabilities almost identical to wild-type and yet had operator binding affinities reduced by 100-fold or more (20, 22, 45). The conclusion that these "activity" mutations identify residues in or near the DNA-recognition surface of the protein has been directly supported by the crystal structure of the protein-DNA complex (28). Similar identification of DNA-binding residues by biochemical characterization of purified mutant proteins has been reported for EcoR1 (77) and P22 Arc repressor (72).

It is sometimes possible to infer that mutant proteins are stably folded without purification and subsequent study. For proteins that are active only as oligomers, stably folded but inactive proteins may have a transdominant negative phenotype because mixed oligomers containing wild-type and mutant subunits have dramatically reduced activities. For example, most dominant-negative mutations in the Trp repressor affect side chains in or near the DNA-binding surface of the protein (32, 63).

With current methods of cassette mutagenesis, functionally important residues can also be identified by studies of neutral mutations. For example, a cassette method was used to mutagenize regions of about 30 base-pairs in the *arc* repressor gene such that most cassettes contained from two to four mutations (8). Following an activity selection, functionally neutral residue substitutions were identified at 24 of the 53 positions of Arc. In a separate screening experiment, mutant Arc sequences that could still fold into a stable structure were isolated, and substitutions, some conservative and some non-conservative, were identified at 41 positions. Comparison of these two sets of neutral mutations revealed that the N-terminal residues of Arc could tolerate substitutions when formation of a stable structure was required but not when function was required, suggesting that these residues form part of the operator-binding surface of the protein. The identification of this region of Arc as the likely DNA-binding region has also been supported by studies of defective mutant proteins (72) and chimeric proteins with Arc-binding specificity (37).

There can clearly be problems in interpretation for any of the experiments discussed above. Some stably folded mutant proteins might have subtly altered conformations that are responsible for their decreased activity; a mutant substitution may exert its main effect directly on activity but also cause a modest decrease in stability. Nevertheless, with appropriate caution, functionally important residues can usually be identified. It is generally easiest to do this when dealing with mutations that cause significant reductions in activity. There are presumably a large number of ways, many of them

subtle, to reduce protein activity by a factor of two. By contrast, there are relatively few ways to reduce activity by a factor of 100-fold or more, and most of these will involve large and easily detectable changes in protein stability or the alteration of functionally important residues.

MUTATIONS THAT ENHANCE STABILITY AND ACTIVITY.

Different genetic strategies have been used to identify mutations that enhance protein stability and/or activity. Most use some means of reducing the activity of the protein of interest, followed by a selection or screen to detect variants with increased activity. For example, the activity of a protein might be reduced by mutation (21, 24, 45, 51), by decreasing its intracellular level, by increasing the temperature (11, 38, 39), or by decreasing the concentration of required cofactors (32). The parental gene is then mutagenized and strains with increased activity can be isolated and analyzed.

In studies in which one starts with a gene bearing a loss of activity mutation, pseudo-revertants can arise at the site of the original mutation or at second-sites within the gene or in other genes (76). The most common types of second-site suppressor mutations are those that act globally to overcome the original defect by increasing protein stability, activity, or level. For example, if a defective mutation destabilizes a protein by 2 kcal/mol, then a second-site substitution might act by increasing stability by a comparable amount. In such a case, an otherwise wild-type protein bearing the suppressor mutation should be more stable than wild-type. Mutations that increase protein stability have been identified in this way for staphylococcal nuclease (65, 67) and λ Cro (51). Enhanced stability mutations have also been identified in kanamycin nucleotidyltransferase and subtilisin by selecting or screening for activity at elevated temperatures (11, 38, 39).

Some second-site suppressor mutations act by increasing activity directly. For example, amino acid substitutions in λ repressor that increase operator-binding affinity as much as 600-fold have been identified by their ability to suppress both stability and activity mutations (21, 45). A similar class of mutations has been identified in Trp repressor, but by direct selection for activity at low concentrations of the co-repressor, tryptophan (32, 36).

PROTEOLYTIC SENSITIVITY OF MUTANT PROTEINS

Since unfolded proteins are usually better substrates for proteolytic digestion than their folded counterparts, intracellular proteolysis of unstable proteins can play an important role in mutant phenotypes. For example, the Ile30→Leu mutation in λ Cro affects a residue in the hydrophobic core and

reduces the T_m of the protein from 40°C to 35°C (51). This decrease in stability alone would only be expected to cause a modest decrease in activity by reducing the concentration of folded, active Cro. However, whereas wild-type Cro has an intracellular half-life of 60 min, the mutant half-life is reduced to 11 min. Hence, proteolysis amplifies the effect of this destabilizing mutation by reducing the steady-state level of the mutant protein.

Several findings suggest that the stability of a folded protein is an important determinant of its rate of degradation. First, proteins that contain amino acid analogs or are prematurely terminated are often degraded rapidly in the cell (16). Second, good correlations exist between the measured or inferred thermal stabilities of specific mutant proteins and the rates at which they are degraded in *E. coli* (52, 53). Finally, second-site suppressor mutations that increase the thermodynamic stability of unstable mutant proteins have also been shown to increase resistance to intracellular proteolysis (51).

The rate of intracellular proteolysis of mutant proteins can also be influenced by determinants other than the stability of the native structure. For example, the N-terminal residues of some proteins appear to be important in determining their susceptibility to ubiquitin-mediated degradation in the yeast *S. cerevisiae* (4). In *E. coli*, the identity of residues at the C-terminal ends of some proteins influences their rates of intracellular degradation (9, 53). For example, frameshift mutations near the C-terminus of the Arc repressor result in the addition of extra C-terminal residues that suppress the proteolytic instability of unstable Arc mutants without affecting the thermal stability or activity of the protein (9). In addition to sequence determinants, the solubility of mutant proteins can also affect their proteolytic resistance. Some proteins aggregate to form inclusion bodies, presumably because they are unfolded or incompletely folded, and thus escape proteolytic attack (29). Because of these factors, increased susceptibility to intracellular degradation does not by itself provide sufficient evidence to conclude that a mutant is thermodynamically unstable. In similar fashion, a mutant protein could be resistant to intracellular proteolysis and yet not be stably folded. Nevertheless, susceptibility to degradation can be a convenient indicator of thermodynamic stability for some proteins.

SUMMARY

There is tremendous variability in the importance of individual amino acids in protein sequences. On the one hand, nonconservative residue substitutions can be tolerated with no loss of activity at many residue positions, especially those exposed on the protein surface. On the other hand, destabilizing mutations can occur at a large number of different sites in a protein, and for many proteins such mutations account for more than half of the randomly isolated

missense mutations that confer a defective phenotype. At sites that are key determinants of stability or activity, even residue substitutions that are generally considered to be conservative (e.g., Glu↔Asp, Asn↔Asp, Ile↔Leu, Lys↔Arg and Ala↔Gly) can have severe phenotypic effects. Unfortunately, this means that there is no simple way to infer the likely effect of an amino acid substitution on the basis of sequence information alone. A nonconservative Gly→Arg substitution could be phenotypically silent at one position while a conservative Asn→Asp change could lead to complete loss of activity at another position.

For proteins whose structures are known, it is often possible to predict whether particular residue substitutions will be destabilizing, as long as detailed estimates of the destabilization energy are not required. Substitutions that introduce polar groups, large cavities, or overly large side chains into the hydrophobic core are potentially the most destabilizing. Substitutions that disrupt hydrogen bonding or electrostatic interactions can also have significant effects, although the destabilization caused by these substitutions is smaller than that caused by severe core mutations. Destabilizing substitutions that involve replacing glycines in turns, or introducing prolines into α-helices and other disallowed positions are also reasonably common. Finally, most solvent exposed residues can apparently be freely substituted without serious effects on protein stability. Although exceptions may occur, these generalizations serve to summarize a large body of information and can be rationalized in physical and chemical terms.

It is an especially encouraging result that proteins appear to tolerate most substitutions, even those that are destabilizing, without significant changes in the native structure. For proteins whose structures are known, this means that it is reasonable to interpret mutant phenotypes in terms of the wild-type structure. For proteins whose structures are not known, it is reasonable to infer that mutations that reduce activity without affecting stability are directly involved in function. Detailed studies of the structure of the mutant proteins are still needed, but, because induced conformational changes are rare, such efforts are usually worthwhile.

Because proteins are so diverse, it is always dangerous to extrapolate too far. We note that most of the studies described here concern small, globular, single-domain proteins whose folded and unfolded structures are in dynamic equilibrium. Fibrous proteins, proteins that are extremely thermostable, or proteins that contain multiple interacting domains may face special problems in folding (34). Moreover, indirect effects of mutations mediated via protein conformation are much more likely to be common for allosteric proteins, which can exist in distinctly different quaternary structures (61). Nevertheless, the basic principles of protein structure and activity established in the simpler and more readily studied systems should still form the groundwork for studies on more complicated proteins.

ACKNOWLEDGMENTS

We thank Jim Bowie and Dawn Parsell for comments on the manuscript, John Reidhaar-Olson for communication of unpublished data, Jim Bowie and Neil Clarke for assistance with solvent accessibility calculations, and Lynn Kleina and Jeffrey Miller for a preprint of their manuscript (35). A.A.P is a postdoctoral fellow of the American Cancer Society, California Division. Work in the author's laboratory was supported by NIH grants AI-15706 and AI-16892 to R.T.S.

Literature Cited

1. Alber, T., Bell, J. A., Sun, D.-P., Nicholson, H., Wozniak, J. A., et al. 1988. Replacements of Pro86 in phage T4 lysozyme extend an alpha-helix but do not alter protein stability. *Science* 239:631–35
2. Alber, T., Sun, D.-P., Nye, J. A., Muchmore, D. C., Matthews, B. W. 1987. Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. *Biochemistry* 26:3754–58
3. Alber, T., Sun, D.-P., Wilson, K., Wozniak, J. A., Cook, S. P., Matthews, B. W. 1987. Contributions of hydrogen bonds of Thr157 to the thermodynamic stability of phage T4 lysozyme. *Nature* 330:41–46
4. Bachmair, A., Finley, D., Varshavsky, A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234:179–86
5. Bartlett, P. A., Marlowe, C. K. 1987. Evaluation of intrinsic binding energy from a hydrogen bonding group in an enzyme inhibitor. *Science* 235:569–71
6. Becktel, W. J., Schellman, J. A. 1987. Protein stability curves. *Biopolymers* 26:1859–77
7. Botstein, D., Shortle, D. 1985. Strategies and applications of in vitro mutagenesis. *Science* 229:1193–201
8. Bowie, J. U., Sauer, R. T. 1989. Identifying determinants of folding and activity for a protein of unknown structure. *Proc. Natl. Acad. Sci. USA* 86: 2152–56
9. Bowie, J. U., Sauer, R. T. 1989. Identification of C-terminal extensions that protect proteins from intracellular proteolysis. *J. Biol. Chem.* 264:7596–602
10. Brown, L. R., DeMarco, A., Richarz, R., Wagner, G., Wuthrich, K. 1978. The influence of a single salt bridge on static and dynamic features of the globular solution conformation of the basic pancreatic trypsin inhibitor. *Eur. J. Biochem.* 88:87–95
11. Bryan, P. N., Rollence, M. L., Pantoliano, M. W., Wood, J., Finzel, B. C., et al. 1986. Proteases of enhanced stability: Characterization of a thermostable variant of subtilisin. *Proteins* 1: 326–34
12. Chou, P. Y., Fasman, G. D. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* 47: 251–76
13. Creighton, T. E. 1984. *Proteins. Structures and Molecular Properties*. New York: Freeman
14. Fauchere, J.-L., Pliska, V. 1983. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem. Chim. Ther.* 18:369–75
15. Fersht, A. R., Shi, J. P., Knill-Jones, L., Lowe, D. M., Wilkinson, A. J., et al. 1985. Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature* 314:235–38
16. Goldberg, A. L., St. John, A. C. 1976. Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annu. Rev. Biochem.* 45:747–803
17. Grütter, M. G., Gray, T. M., Weaver, L. H., Alber, T., Wilson, K., Matthews, B. W. 1987. Structural studies of mutants of the lysozyme of bacteriophage T4: The temperature-sensitive mutant protein Thr157→Ile. *J. Mol. Biol.* 197:315–29
18. Hampsey, D. M., Das, G., Sherman, F. 1986. Amino acid replacements in yeast iso-1-cytochrome *c*: comparison with the phylogenetic series and the tertiary structure of related cytochromes-*c*. *J. Biol. Chem.* 261:3259–71
19. Hampsey, D. M., Das, G., Sherman, F.

1988. Yeast iso-1-cytochrome *c*: genetic analysis of structural requirements. *FEBS Lett.* 231:275–83
20. Hecht, M. H., Nelson, H. C. M., Sauer, R. T. 1983. Mutations in lambda repressor's amino-terminal domain: Implications for protein stability and DNA binding. *Proc. Natl. Acad. Sci. USA* 80:2676–80
21. Hecht, M. H., Sauer, R. T. 1985. Phage lambda repressor revertants: Amino acid substitutions that restore activity to mutant proteins. *J. Mol. Biol.* 186:53–63
22. Hecht, M. H., Sturtevant, J. M., Sauer, R. T. 1984. Effect of single amino acid replacements on the thermal stability of the NH_2-terminal domain of phage lambda repressor. *Proc. Natl. Acad. Sci. USA* 81:5685–89
23. Hecht, M. H., Sturtevant, J. M., Sauer, R. T. 1986. Stabilization of lambda repressor against thermal denaturation by site-directed Gly→Ala changes in alpha-helix 3. *Proteins* 1:43–46
24. Ho, C., Jasin, M., Schimmel, P. 1985. Amino acid replacements that compensate for a large polypeptide deletion in an enzyme. *Science* 229:389–93
25. Hol, W. G., Halie, L. M., Sander, C. 1981. Dipoles of the alpha-helix and beta-sheet and their role in protein folding. *Nature* 294:532–36
26. Hope, I. A., Struhl, K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of Yeast. *Cell* 46:885–94
27. Jasin, M., Regan, L., Schimmel, P. 1983. Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. *Nature* 306:441–47
28. Jordan, S. R., Pabo, C. O. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* 242:893–99
29. Kane, J. F., Hartley, D. L. 1988. Formation of recombinant protein inclusion bodies in Escherichia coli. *Tibtech* 6:95–101
30. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63
31. Keegan, L., Gill, G., Ptashne, M. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* 231:699–704
32. Kelley, R. L., Yanofsky, C. 1985. Mutational studies with the trp repressor of *Escherichia coli* support the helix-turn-helix model of repressor recognition of operator DNA. *Proc. Natl. Acad. Sci. USA* 82:483–87
33. Kellis, J. T. J., Nyberg, K., Sali, D., Fersht, A. R. 1988. Contribution of hydrophobic interactions to protein stability. *Nature* 333:784–86
34. King, J. 1986. Genetic analysis of protein folding pathways. *BioTechnology* 4:297–303
35. Kleina, L. G., Miller, J. H. 1989. Genetic studies of the *lac* repressor: XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *J. Mol. Biol.* Submitted
36. Klig, L., Yanofsky, C. 1988. Increased binding of operator DNA by trp superrepressor EK49. *J. Biol. Chem.* 263:243–46
37. Knight, K. L., Sauer, R. T. 1989. DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues. *Proc. Natl. Acad. Sci. USA* 86:797–801
38. Liao, H., McKenzie, T., Hageman, R. 1986. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc. Natl. Acad. Sci. USA* 83:576–80
39. Matsumura, M., Aiba, S. 1985. Screening for thermostable mutants of kanamycin nucleotidyltransferase by the use of a transformation system for a thermophile, *Bacillus stearothermophilus*. *J. Biol. Chem.* 260:15298–303
40. Matsumura, M., Becktel, W. J., Matthews, B. W. 1988. Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* 334:406–10
41. Matsumura, M., Yahanda, S., Yasumura, S., Yutani, K., Aiba, S. 1988. Role of tyrosine-80 in the stability of kanamycin nucleotidyltransferase analyzed by site-directed mutagenesis. *Eur. J. Biochem.* 171:715–20
42. Matthews, B. W. 1987. Genetic and structural analysis of the protein stability problem. *Biochemistry* 26:6885–88
43. Matthews, B. W., Nicholson, H., Becktel, W. J. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA* 84:6663–67
44. Miller, J. H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H., et al. 1979. Genetic studies of the *lac* repressor: ix. generation of altered proteins by the suppression of nonsense mutations. *J. Mol. Biol.* 131:191–222
45. Nelson, H. C., Sauer, R. T. 1985. Lambda repressor mutations that in-

crease the affinity and specificity of operator binding. *Cell* 42:549–58
46. Nicholson, H., Becktel, W. J., Matthews, B. W. 1988. Enhanced protein thermostability from designed mutations that interact with α-helix dipoles. *Nature* 336:651–56
47. Normanly, J., Abelson, J. 1989. tRNA identity. *Annu. Rev. Biochem.* 58:1029–49
48. Nozaki, Y., Tanford, C. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* 246: 2211–17
49. Pace, C. N. 1975. The stability of globular proteins. *CRC Crit. Rev. Biochem.* 3:1–43
50. Pace, C. N., Barrett, A. J. 1984. Kinetics of tryptic hydrolysis of the arginine-valine bond in folded and unfolded ribonuclease T1. *Biochem. J.* 219:411–17
51. Pakula, A. A., Sauer, R. T. 1989. Amino acid substitutions that increase the thermal stability of the λ Cro protein. *Proteins* 5:202–10
52. Pakula, A. A., Young, V. B., Sauer, R. T. 1986. Bacteriophage lambda cro mutations: Effects on activity and intracellular degradation. *Proc. Natl. Acad. Sci. USA* 83:8829–33
53. Parsell, D. A., Sauer, R. T. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *E. coli*. *J. Biol. Chem.* 264:7590–95
54. Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., et al. 1987. Effect of single amino acid replacements on the folding and stability of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* 26:2674–82
55. Ramdas, L., Sherman, F., Nall, B. T. 1986. Guanidine hydrochloride induced unfolding of mutant forms of iso-1-cytochrome *c* with replacement of proline-71. *Biochemistry* 25:6952–58
56. Reidhaar-Olson, J. F., Sauer, R. T. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241: 53–57
57. Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Biophys. Chem.* 6:151–76
58. Richardson, J. S., Richardson, D. C. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240:1648–52
59. Rose, G. D., Gierasch, L. M., Smith, J. A. 1985. Turns in peptides and proteins. *Adv. Protein Chem.* 37:1–109
60. Sali, D., Bycroft, M., Fersht, A. R. 1988. Stabilization of protein structure by interaction of alpha-helix dipole with a charged side chain. *Nature* 335:740–43
61. Schachman, H. K. 1988. Can a simple model account for the allosteric transition of aspartate transcarbomoylase? *J. Biol. Chem.* 263:18583–86
62. Schellman, J. A., Lindorfer, M., Hawkes, R., Grutter, M. 1981. Mutations and protein stability. *Biopolymers* 20:1989–99
63. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., Sigler, P. B. 1985. The three-dimensional structure of *trp* repressor. *Nature* 317: 782–86
64. Schimmel, P. R., Flory, P. J. 1968. Conformational energies and configurational statistics of copolypeptides containing L proline. *J. Mol. Biol.* 34: 105–20
65. Shortle, D. 1986. Guanidine hydrochloride denaturation studies of mutant forms of staphylococcal nuclease. *J. Cell. Biochem.* 30:281–89
66. Shortle, D., DiMaio, D., Nathans, D. 1981. Directed mutagenesis. *Annu. Rev. Genet.* 15:265–94
67. Shortle, D., Lin, B. 1985. Genetic analysis of staphylococcal nuclease: Identification of three intragenic "global" suppressors of nuclease-minus mutations. *Genetics* 110:539–55
68. Shortle, D., Meeker, A. K. 1986. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins* 1:81–89
69. Shortle, D., Meeker, A. K. 1989. Residual structure in large fragments of staphylococcal nuclease: Effects of amino acid substitutions. *Biochemistry* 28: 936–44
70. Shortle, D., Meeker, A. K., Freire, E. 1988. Stability mutants of staphylococcal nuclease: Large compensating enthalpy-entropy changes for the reversible denaturation reaction. *Biochemistry* 27: 4761–68
71. Tronrud, D. E., Holden, H. M., Matthews, B. W. 1987. Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond. *Science* 235:571–74
72. Vershon, A. K., Bowie, J. U., Karplus, T. M., Sauer, R. T. 1986. Isolation and analysis of arc repressor mutants: evidence for an unusual mechanism of DNA binding. *Proteins* 1:302–11
73. Wells, J. A., Powers, D. B. 1986. In vivo formation and stability of engi-

neered disulfide bonds in subtilisin. *J. Biol. Chem.* 261:6564–70
74. Wetzel, R., Perry, L. J., Baase, W. A., Becktel, W. J. 1988. Disulfide bonds and thermal stability in T4 lysozyme. *Proc. Natl. Acad. Sci. USA* 85:401–5
75. Wilde, J. A., Bolton, P. H., Dell' Acqua, M., Hibler, D. W. Pourmotabbed, T., Gerlt, J. A. 1988. Identification of residues involved in a conformational change accompanying substitutions for glutamate-43 in staphylococcal nuclease. *Biochemistry* 27:4127–32
76. Yanofsky, C. 1971. Tryptophan biosynthesis in *Escherichia coli*. Genetic determination of the proteins involved. *J. Am. Med. Assoc.* 218:1026–35
77. Yanofsky, S. D., Love, R., McClarin, J. A., Rosenberg, J. M., Boyer, H. W., Greene, P. J. 1987. Clustering of null mutations in the EcoRI endonuclease. *Proteins* 2:273–82
78. Yutani, K., Ogasahara, K., Tsujita, T., Sugino, Y. 1987. Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl. Acad. Sci. USA* 84: 4441 44
79. Zoller, M. J., Smith, M. 1987. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *Methods Enzymol.* 154:329–50

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☑ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.